Passive Acoustic Detection and Real-Time Localization of Belugas in the St. Lawrence Estuary

Clément Chion, Emmanuel Fernandez, Paul Ducrocq, Paul Laurent, Dominic Lagrois, Irene Roca, Jean-François Senécal, and Robert Michaud

Contents

OPEN ACCESS with major support from JASCO APPLIED SCIENCES

C. Chion (✉) · E. Fernandez · P. Ducrocq · P. Laurent · D. Lagrois · I. Roca · J.-F. Senécal
Department of Natural Sciences, Université du Québec en Outaouais, Gatineau, QC, Canada
e-mail: clement.chion@uqo.ca; paul.laurent@etu.minesparis.psl.eu; dominic.lagrois.1@ulaval.ca; irene.rocatorrecilla@uqo.ca

R. Michaud
Groupe de Recherche et d' Éducation sur les Mammifères Marins, Tadoussac, QC, Canada
e-mail: rmichaud@gremm.org

A. N. Popper et al. (eds.), *The Effects of Noise on Aquatic Life IV*,
https://doi.org/10.1007/978-3-031-94229-7_171-1

Abstract

Beluga whales (*Delphinapterus leucas*) rely extensively on acoustic signaling for communication, navigation, foraging, and social interactions. Their vocal repertoire spans a broad frequency range from several hundred hertz to approximately 160 kHz. Consequently, passive acoustic monitoring (PAM) provides a robust and reliable method for detecting their presence, with a low false-negative rate. The endangered St. Lawrence Estuary Beluga population faces multiple anthropogenic stressors that hinder its recovery, among which underwater noise from marine traffic ranks among the top three threats. Given that the St. Lawrence River serves as a major global commercial waterway, real-time detection, local abundance estimation, and localization of belugas offer substantial conservation benefits. One key application is the ability to alert approaching vessels, thus minimizing close encounters and reducing noise exposure. This chapter introduces three integrated modules of a developing PAM-based system that combines ground truth datasets with artificial intelligence techniques to enable automated detection of beluga presence, estimation of herd size, and real-time localization of individuals, supporting both conservation management and mitigation of acoustic disturbances.

Keywords

St. Lawrence Estuary beluga whale · Passive acoustic monitoring · Artificial intelligence · Real-time automated system · Marine traffic noise · Conservation

Introduction

The beluga whale (*Delphinapterus leucas*) population inhabiting the St. Lawrence Estuary (SLEB) represents an isolated group that has been designated as endangered since 2014 (COSEWIC 2014). The population size is currently estimated at about 1800 individuals, which still falls below the Precautionary Reference Level of 3219 individuals (DFO 2023). Multiple anthropogenic stressors have been identified as key threats impeding population recovery, such as, chronic exposure to industrial contaminants, reduced prey availability, vessel traffic, and anthropogenic noise (Lesage 2021). In particular, the pervasive exposure to the radiated noise from marine traffic has been related to spatial displacements (Small et al. 2017), behavioral changes (Martin et al. 2023) and significant reductions in the acoustic communication space of beluga whales (Vergara et al. 2021). However, quantifying the impacts of marine traffic on SLEB habitat use, fundamental activities and population dynamics remains challenging, due to the difficulty of accessing the animals in their natural environment and their endangered status.

A rapid and systematic reduction of vessel traffic and/or the associated underwater radiated noise in the St. Lawrence Estuary (SLE) remains a complex issue, largely a consequence of significant socio-economic constraints. On the one hand, while technology development to make vessels quieter (elevating engines off the ship floor, hull/propeller design) is ongoing, there is an inevitable implementation lag. On the other hand, the St. Lawrence River is a major global commercial waterway, and several economically relevant seafaring activities also contribute to elevated underwater noise levels in the beluga's critical habitat in the SLE. Commercial shipping and regular ferry operations contribute with chronic year-round inputs, while the mature and flourishing whale-watching industry and recreational boating are responsible for additional seasonal peaks (Gervaise et al. 2012). Therefore, action plans to reduce noise in the SLEB habitat need forcibly to represent a compromise between conservation goals and socioeconomic needs. There are currently some integrative, government-led action plans in the SLE to reduce the impact of vessel noise on SLEB, such as the Action Plan to Reduce the Impact of Noise on the Beluga Whale and Other Marine Mammals at Risk in the St. Lawrence Estuary (DFO 2020) or the marine activities regulations at the Saguenay-St. Lawrence Marine Park (e.g., Chion et al. 2018). Nevertheless, most of the practical measures that are currently implemented, including speed reductions and re-routing, are static (activated or deactivated during particular periods and covering specific areas) and voluntary, with varying public and industry compliance. Besides, there are still areas of SLEB critical habitat with insufficient monitoring coverage or enforcement.

Recent technological developments are promoting the realistic implementation of dynamic monitoring tools with integrated rapid and automated analysis of passive acoustic data. These systems can provide real-time information about whales' presence and location, as well as alert nearby vessels for an eventual change in speed, bearing or track. Examples are the NOAA Right Whale Listening Network (Hyer et al. 2025) or the BOMBYX project (Glotin et al. 2021). Passive acoustic monitoring (PAM) constitutes a non-invasive, weather-independent approach that enables continuous recording to investigate marine mammals' habitat use and behavior, being particularly useful in remote or logistically challenging environments (e.g., Roca et al. 2023). Beluga whales exhibit a highly diverse acoustic repertoire and consistently employ acoustic signals for communication, navigation, foraging, and social interactions (Sjare and Smith 1986; Vergara and Mikus 2019; Vergara et al. 2025). Therefore, PAM coupled with advanced machine learning techniques, may also have a great potential to serve as a means of developing dynamic conservation tools for SLEB.

This chapter presents three modules of an automated system: an integrative pipeline for real-time SLEB call-type detection and classification, and two downstream analytical components—estimation of local population abundance from vocal activity time series and individual localization using time-difference-of-arrival across hydrophone arrays. The main goal of this research is to develop and apply an automated system based on the combination of PAM data and Artificial Intelligence (AI), for real-time monitoring of SLEB to eventually alert vessels of their presence in high-traffic areas of the SLE.

Call Classification Pipeline

A real-time acoustic pipeline was implemented to detect and classify SLEB's vocalizations. Full technical details are provided in Cotillard et al. (2024) and Fernandez et al. (2025); below the authors summarize the key elements relevant to this chapter.

Acoustic Data

The model was trained on acoustic recordings from four sites in the SLE system: Baie Sainte-Marguerite (BSM), Cacouna (CAC), Rivière-du-Loup (RDL), and Kamouraska (KAM) (Fig. 1). Site selection was informed by existing knowledge of regions within the SLE known to play a key role in SLEB ecology. Acoustic recordings were collected with SoundTrap hydrophones (Ocean Instruments, NZ) (see Fernandez et al. 2025, Table A1 for details). SLEB's acoustic signals were manually labelled by expert bioacousticians by visually and aurally inspecting spectrograms from the recordings. Acoustic signals were labelled into four distinct and comprehensive call types which represent the complete vocal repertoire of SLEB (Fig. 2):

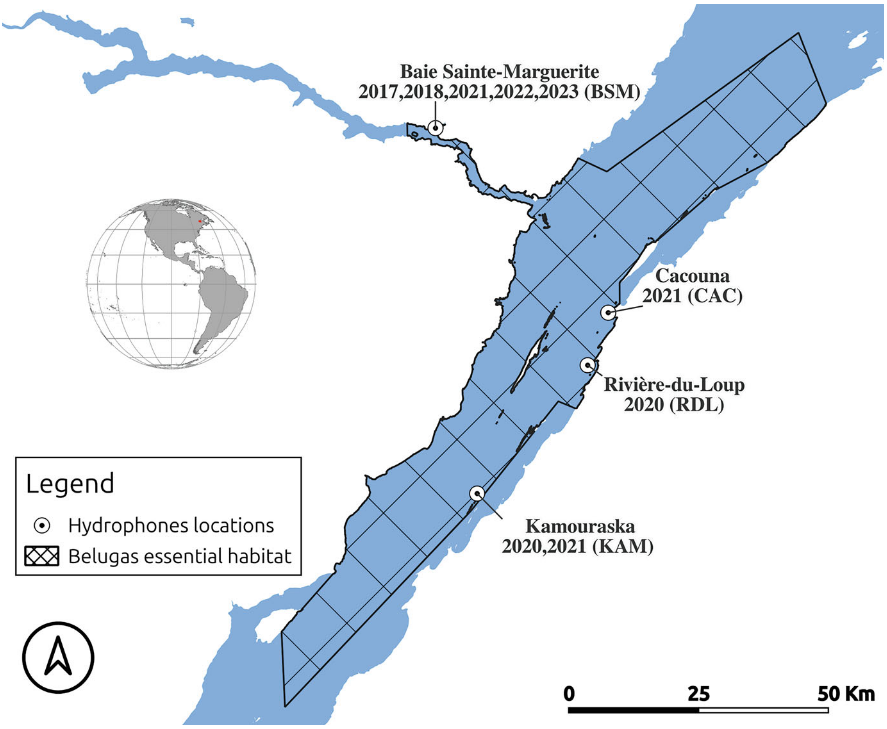


Fig. 1 Map of the study area showing the sampling sites

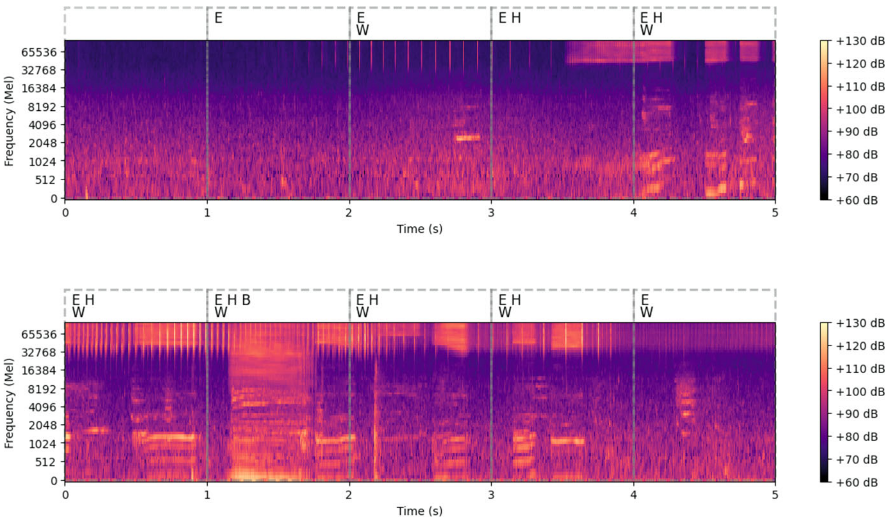


Fig. 2 Example outputs of the call classification pipeline, with one prediction per second on a recording from BSM (July 13, 2018). Detected call types are overlaid above the spectrogram: E for ECHO, H for HFPC, B for BBPC, and W for Whistles

- Echolocation Clicks (ECHO): Broadband (~30–150 kHz) pulsed signals used primarily for navigation and prey localization (Sjare and Smith 1986). In this study, ECHOs denote echolocation events consisting of multiple pulses within 1 s recording.
- High-Frequency Pulsed Calls (HFPC): Pulsed bursts of ultrasonic broadband sounds (>40 kHz). HFPCs can be distinguished from burst pulses in echolocation buzz trains by their high and constant pulse repetition rate throughout the signal (Vergara et al. 2025).
- Broadband Pulsed Calls (BBPC): Broadband signals (200 Hz–144 kHz) with substantial acoustic energy. The most studied subcategory is the contact call, a long-duration (typically >1 s) pulsed signal used for group cohesion, such as mother–calf interactions (Vergara and Mikus 2019).
- Whistles (W): Low-frequency (~200 Hz to 20 kHz) narrowband modulated tones or pulsed vocalizations. Whistles are commonly associated with social communication (Sjare and Smith 1986).

CNN Model

The classification model was based on a single, optimized convolutional neural network (CNN). It processes 1-second spectrogram windows and outputs independent binary predictions for each studied call type (multi-label setup), enabling robust detection of overlapping vocalizations. The model attained 94% mean classification accuracy on held-out data. After quantization, it had an approximately 320 KB footprint suitable for embedded/edge deployment in dynamic monitoring systems (Fernandez et al. 2025).

Pipeline Implementation

The pipeline operates by iteratively processing audio in 1-s non-overlapping windows using the model, as illustrated in Fig. 2. The fixed 1-s window length imposes a limitation on the precise counting of calls. For instance, a single long-duration call spanning over two windows may be counted twice, whereas multiple short calls occurring within one window would yield a single positive detection. However, in practice, the effect is largely systematic; the resulting time series provide a reliable proxy of relative call abundance and vocal activity at the temporal scales used in the downstream localization and abundance modules.

Local Abundance Estimation

Automated estimation of local SLEB abundance from underwater vocalizations can improve the understanding of spatial habitat use in the SLE. If available in near-real time, this ecological information could also guide dynamic management of vessel routes and speeds. In the system's second module, the authors tested whether call activity detected by the classification pipeline could act as a proxy for local herd size. To do so, they developed a two-stage machine-learning approach that first detects beluga presence, then estimates the number of individuals within the hydrophone's listening range.

Beluga Count Data

The study area was located in the Saguenay Fjord section of the Saguenay–St. Lawrence Marine Park, at Baie Sainte-Marguerite (BSM) (cf. Fig. 1.). Training data comprised paired observations from BSM, where systematic land-based visual counts were temporally aligned with co-located hydrophone recordings for the summer seasons of 2018, 2021, 2022, and 2023. Land-based observations were conducted from the "Halte du Béluga" lookout, part of the Fjord-du-Saguenay National Park (Sépaq), situated 16 meters above sea level on the northern shore of the fjord.

Land-based observations normally take place between mid-June and mid-September in the Marine Park. The observer is stationed at a fixed viewpoint and uses Bushnell 7 × 50 binoculars equipped with reticle scales and an integrated compass. The binoculars are mounted on a tripod and set up each day in the same location with consistent settings, using established terrestrial reference points to ensure repeatability across observers. Belugas are counted during fixed-length observation blocks, which lasted 40 minutes in 2018 and were reduced to 15 min in subsequent years. To ensure coverage across the entire daylight period (7:00 to 19:00), observers vary the sequence of observation blocks and breaks each day. The observer estimates the number of individuals present as accurately as possible, considering diving behavior; rather than recording only the number of animals visible at a specific moment, the observer provides an integrated estimate of group size over the observation period. In addition to group size estimates, contextual variables such as visibility, Beaufort Sea state, confidence in estimate and other observational conditions are also recorded through written comments. Although each observation is conducted as accurately as possible, uncertainty may still arise, particularly in cases involving large groups or when animals surface infrequently. In such cases, the observer may refine the estimate in subsequent observation periods. To account for this, each observation season is followed by a manual review process. A trained reviewer identifies time periods of stable group composition, where no new individuals entered or exited the observation zone, and selects the best estimate of group size based on a combination of raw counts and observer notes.

Classification and Regression Models

The task was framed as a two-stage cascade: a classifier first detected beluga presence, and a regressor then estimated group size. Both stages used Random Forests, with the classifier distinguishing presence from absence under variable acoustic conditions and the regressor modelling the nonlinear link between calling activity and group size.

Only outputs from the call-classification pipeline, aggregated into non-overlapping 10-min windows, were used as model inputs. For each window, features included counts of each call type and moving averages over 30 min, 1 hour, 2 hours, and 6 hours to capture short- and medium-term trends. This design preserves information from periods when large groups are intermittently silent. To avoid data leakage and ensure robust evaluation, cross-validation was performed using years as folds: with 4 years of data, the model was trained on 3 years and tested on the remaining one in turn.

For the detection task, performance was assessed using standard binary classification metrics: accuracy, precision, recall, and F1-score (harmonic mean of precision and recall, balancing both error types). For the regression task, the coefficient of determination (R^2, measuring the proportion of variance explained by the model), mean absolute error (MAE, average absolute difference between predicted and observed counts), and mean absolute percentage error (MAPE, relative error expressed as a percentage) were reported.

Results

For the detection task, the RF classifier achieved an overall accuracy of 0.93 (weighted F1 = 0.93). As shown in Fig. 3, detection of presence was highly reliable (precision = 0.94, recall = 0.97, F1 = 0.96), while absence proved more challenging (precision = 0.89, recall = 0.77, F1 = 0.83).

For the regression task, using the median prediction across observations yielded an average $R^2 = 0.21$, MAE = 12.8, and MAPE = 0.54%. Predictions aligned

Fig. 3 Confusion matrix showing the performance of the detection model for classifying beluga presence and absence

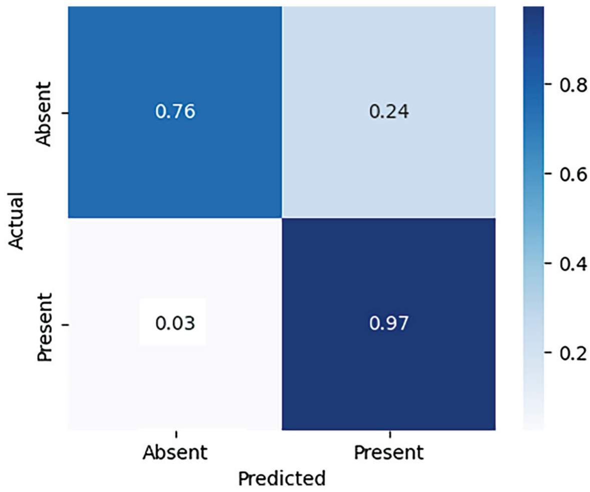
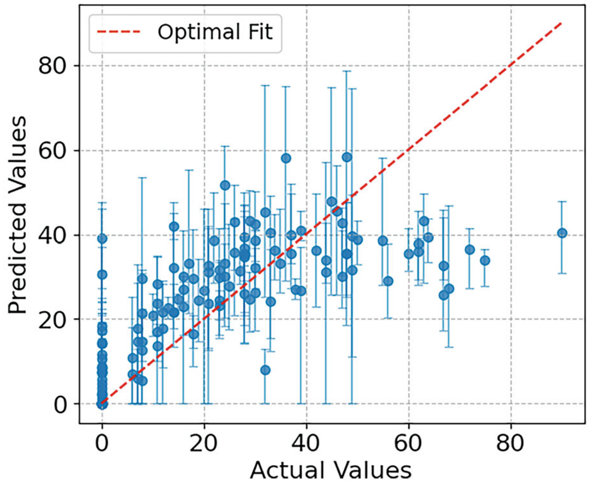


Fig. 4 Predicted versus observed beluga counts for each observation. Points represent mean model predictions, with vertical bars showing the range of predicted values. The red dashed line denotes the 1:1 optimal fit



reasonably well with observed values in the 20–50 range but consistently underestimated group sizes above 50 individuals (Fig. 4).

Discussion

The RF classifier was highly effective in identifying periods of beluga physical presence based on acoustic recordings. The model only erred 0.03% of the time. This result agrees with authors' expectations, given that belugas are highly vocal animals, and highlights the prominence of vocalization within the beluga's behavioral repertoire at BSM. Silent periods in the presence of whales are rare.

The model performed with lower accuracy at correctly predicting absences. Since features were derived exclusively from vocal activity, this outcome reflects cases where the ground-truth label indicated zero belugas, yet calls were nonetheless detected. There may be two main reasons for this. First, false positives from the call-classification pipeline can propagate to the detection model. Although the pipeline achieves an F1 score of 0.93 on the studied call types, no system is error-free, and such misclassifications inevitably affect downstream tasks. Ongoing evaluation of the pipeline in different acoustic environments will be important for characterizing its robustness and limitations. Second, some detections may be genuine calls produced by belugas, either not detected by the observer or outside the visual survey area but still within acoustic range of the hydrophone. At BSM, the semi-enclosed nature of the fjord likely reduces long-range propagation, but through manual spectrogram verification such external calls were confirmed. Without precise measurements of the hydrophone's detection radius, some level of mismatch between acoustic and visual data is unavoidable.

The regression model showed a weak performance at predicting SLEB local abundance solely based on different call type counts. Several factors may explain this result. From an ecological perspective, beluga vocal behavior is not linearly related to group size. For instance, small groups engaged in socializing may be very vocal, while larger groups resting may produce few calls. These behavioral dynamics complicate the direct relationship between vocal activity and abundance. Methodologically, the current approach aggregates call counts over 10-min windows and applies moving averages. While this captures overall activity levels, it discards fine-scale temporal patterns that may contain valuable information about group behavior which could improve the abundance estimation. Incorporating such dynamics would likely require more advanced models (e.g., recurrent or sequence-based deep learning), but the current dataset is not sufficiently large to support such approaches.

Finally, label uncertainty must be considered. Counting exact numbers of belugas in a large area is inherently difficult. Asynchronous surfacing of individuals from a same group, large herds with more or less dispersed groups, or adverse weather conditions (direct or reflected sunlight, mist) render the task daunting. Observers have noted that large group sizes are often underestimated. Such labelling noise introduces additional variability into the regression task and likely constrains overall performance.

The BSM site offered an ideal testbed by pairing systematic visual counts with co-located PAM recordings, but three key challenges emerged: (i) natural variability in beluga vocal behavior, (ii) uncertainty in the hydrophone's detection radius relative to the visual survey area, and (iii) the difficulty of obtaining precise observer counts. These issues are likely to be even greater at other SLEB sites, many of which resemble open-water conditions that complicate fine-scale abundance estimation. In such areas, localization-based methods may provide more reliable minimum counts (e.g., Spiesberger et al. 2021). Future work will therefore examine which call types or temporal patterns best predict group size, to inform more robust abundance estimators and clarify when and where such systems can most effectively support conservation and mitigation efforts.

TDOA-Informed Localization

Traditional software tools, such as PamGuard (Gillespie et al. 2008), have made real-time detection and tracking of marine mammals possible using time-difference-of-arrival (TDOA) methods. However, they are not optimized for the beluga highly diverse vocal repertoire, which spans an unusually wide frequency range (0.5–150 kHz) (see the "Call Classification Pipeline" section above for details on beluga's vocal repertoire). Moreover, many existing tools—such as real-time passive acoustic monitoring systems for dynamic management or offline detectors used for call classification and bearing estimation—focus primarily on detecting calls or estimating bearings. However, they often lack robust three-dimensional localization capabilities and rigorous estimates of positional uncertainty (e.g., Cario et al. 2021). As the third module of the automated system, an algorithm was developed designed to estimate SLEB positions based on the TDOA of their acoustic signals to hydrophone tetrahedra. The algorithm was tested with synthetic vocalizations and in situ deployments.

Localization Algorithm

Among the call types classified by the pipeline, whistles and HFPCs were prioritized to use for localization, since they propagate well and can be captured simultaneously at two arrays. For each call candidate, TDOAs between hydrophones were estimated using cross-correlation (Gillespie et al. 2020). When a beluga vocalization arrives at two hydrophones separated by a distance d, the relative delay (τ) is proportional to the incident angle of the wavefront (Fig. 5 Left).

To improve the fidelity of these TDOA estimates in noisy and multipath-prone aquatic environments, the authors applied a suite of refinements—including band-

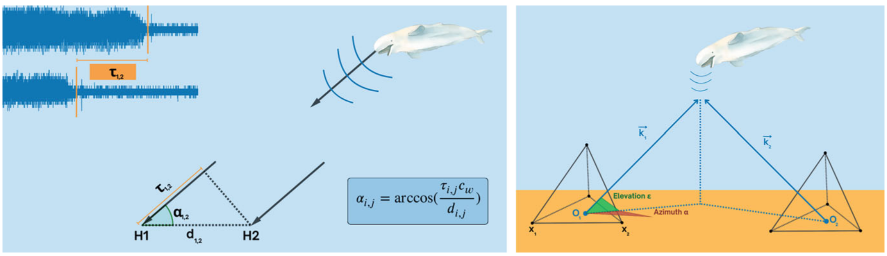


Fig. 5 TDOA and localization geometry. (Left) Principle of TDOA: a beluga call arrives at two hydrophones with a time delay τ. (Right) Geometric representation of two tetrahedral arrays used to localize a beluga vocalization in 3D space

pass filtering, parabolic interpolation around the correlation peak, imposition of maximum delay bounds, and rejection of low-quality or outlier correlations. These measures reduce bias in peak selection, limit physiologically implausible delays, and filter unreliable measurements, thereby bolstering the robustness of downstream localization.

Each tetrahedral array of four hydrophones generates multiple TDOAs. These can be combined into a wave vector pointing toward the sound source. With two arrays, the intersection of these vectors yields a 3D source position (Fig. 5 Right). To improve robustness, a weighted least-squares "low-level fusion" that integrates all TDOAs simultaneously was implemented. This approach avoids discarding partial information and ensures scalability if more arrays are added.

A distinctive feature of the algorithm is its explicit error modeling. Theoretical Cramér–Rao and Ziv–Zakai bounds (CRLB and ZZB) were computed to predict how the variance of TDOA estimates changes as a function of signal-to-noise ratio (SNR), center frequency, and signal bandwidth (Chan and Ho 1994; Bell et al. 1994). These analytical bounds allowed to distinguish three performance regimes: in the random regime (low SNR), TDOA estimates are dominated by noise and highly variable; in the ambiguous regime (intermediate SNR), multiple correlation peaks can compete, leading to increased uncertainty and possible mis-assignment; and in the resolved regime (high SNR), the true delay peak is consistently identified, and variance approaches the theoretical minimum given by the CRLB.

The overall system, from the pipeline for signal detection and classification to the TDOA estimation and data fusion to render an SLEB 3D localization is summarized in Figs. 6 and 7.

Experimental Design

Two custom tetrahedral tripods were constructed with SoundTrap ST640HF four-channel recorders and deployed between 2024 July 26 and August 5 off the city of Cacouna in the SLE south shore, at 20 m depth (Fig. 1). Each tripod carried four HTI hydrophones in a tetrahedral geometry, with the arrays separated by ~500 m. Controlled vessel passages with GPS-tracked trajectories were conducted to support calibration of tripod orientation. Drone flights occurring simultaneously to acoustic recording allowed for ground-truth SLEB herds positioning within the area.

Results

Monte Carlo simulations, in which synthetic whistles were embedded in real background noise, confirmed the realism of the original predictions. At the localization stage, the spatial distribution of errors also closely matched the theoretical model, with only a slight reduction in precision observed between the two tetrahedral arrays.

The 10-day Cacouna deployment yielded both vessel passages and beluga detections. Acoustic reconstructions of vessel tracks aligned with GPS trajectories once

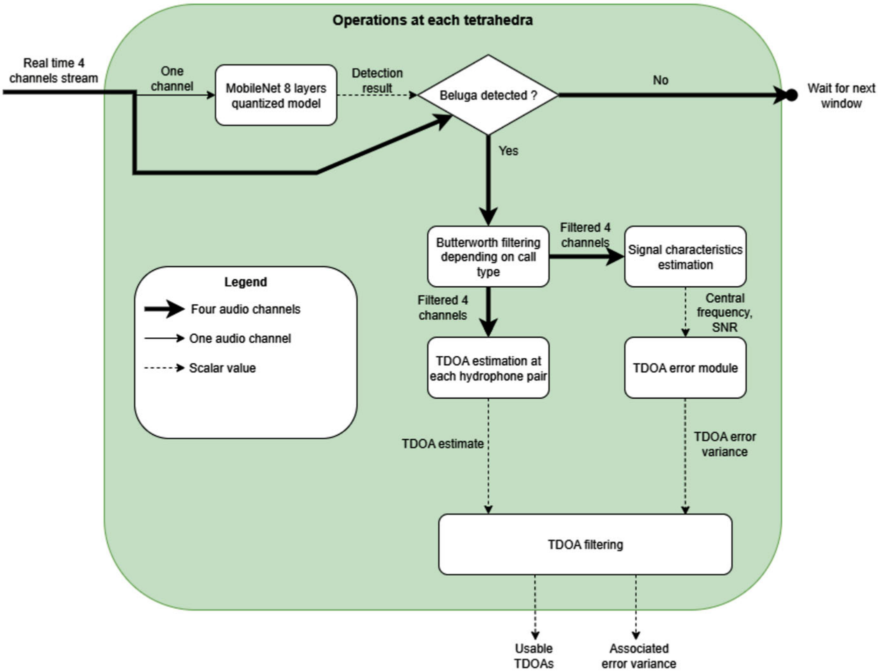


Fig. 6 Block diagram of operations at each tetrahedron, including detection, filtering, TDOA estimation, and error variance computation

tripod orientations had been calibrated. Beluga whistles were localized a posteriori after instrument recovery (Fig. 8). The estimated positions were not always consistent with drone observations, which is not necessarily a limitation of the algorithm: surface observations cannot indicate which individual within a group was producing vocalizations at a given time (see "Discussion").

The TDOA algorithm processed detections within seconds, enabling near real-time localization in the future. The analysis further revealed that the SNR threshold at which errors transition from ambiguous to resolved is frequency-dependent: lower-frequency whistles required higher SNR to achieve accurate localization, while higher-frequency calls were more robust. The fusion module combines all TDOAs and their variances in a covariance-weighted least-squares matrix, producing both a source position and an uncertainty ellipsoid. This emphasizes that localization is not a single point estimate but a probabilistic distribution.

Discussion

This work demonstrates the feasibility of real-time localization of belugas with quantified uncertainty. Compared with generic tools like PamGuard (Gillespie et al. 2008), this system is optimized for beluga call characteristics and integrates

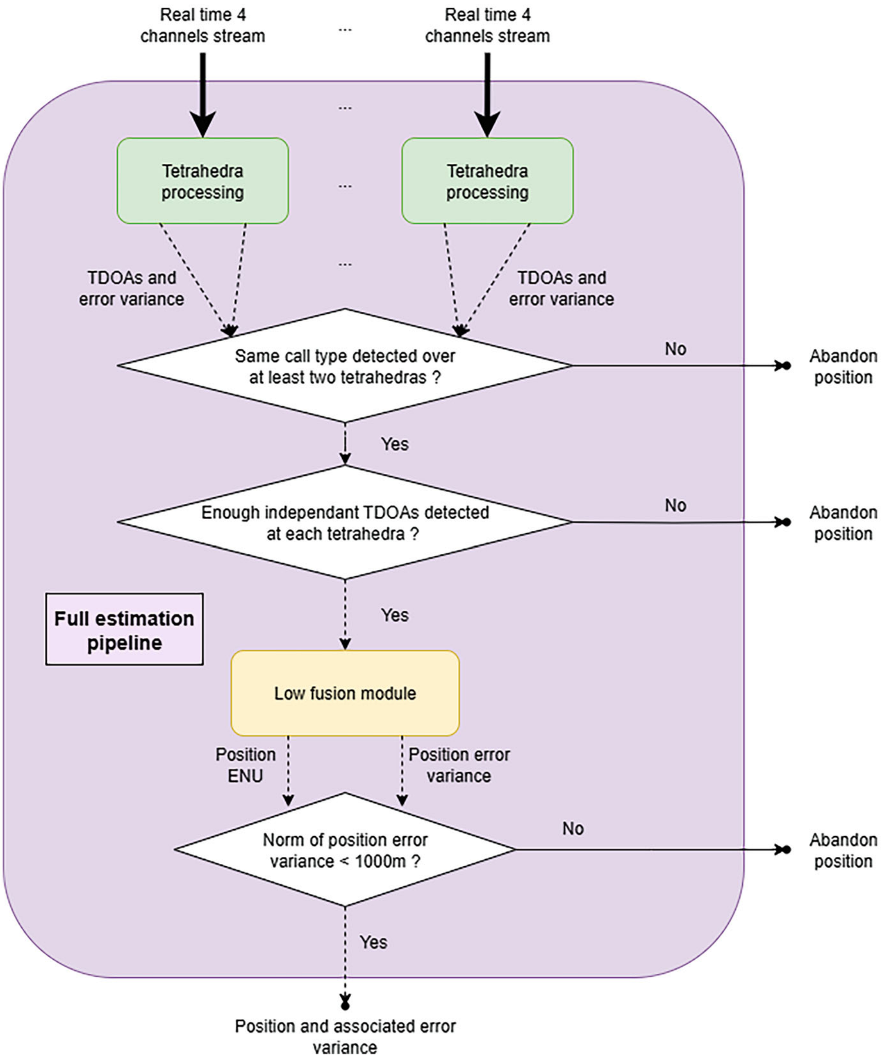
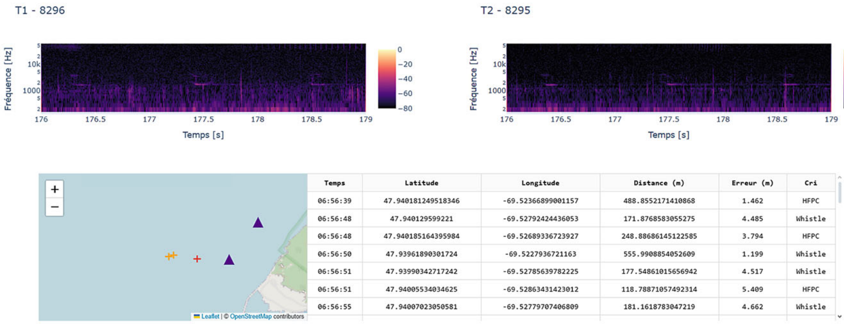


Fig. 7 Global pipeline diagram showing how outputs from multiple tetrahedra are combined to generate 3D positions, including the covariance-weighted fusion step

error bounds. Thus far, the algorithm has been evaluated using synthetic data and during in situ deployments. To address the limitation between 3D positioning and drone observations, a controlled playback experiment with an underwater loudspeaker is planned for the coming weeks, offering true references for assessing whether the algorithm reproduces known source positions, whether predicted uncertainties reflect actual in-situ errors, and what the effective detection range is at this site.



Temps	Latitude	Longitude	Distance (m)	Erreur (m)	Cri
06:56:39	47.940181249518346	-69.52366899001157	488.8552171410868	1.462	HFPC
06:56:48	47.940129599221	-69.52792424436053	171.8768583055275	4.485	Whistle
06:56:48	47.940185164395984	-69.52689336723927	248.88686145122585	3.794	HFPC
06:56:50	47.93961890301724	-69.5227936721163	555.9908854052609	1.199	Whistle
06:56:51	47.93990342717242	-69.52785639782225	177.54861015656942	4.517	Whistle
06:56:51	47.94005534034625	-69.52863431423012	118.78871057492314	5.409	HFPC
06:56:55	47.94007023050581	-69.52779707406809	181.1618783047219	4.662	Whistle

Fig. 8 Interface for beluga detection and localization at Cacouna. Top: spectrograms from both tetrahedra. Bottom: map of array positions (triangles), ground-truth animal positions from drone observations (gold crosses), and estimated positions from the algorithm (red cross). The table summarizes detection times, estimated positions, errors, and call types

Challenges remain, particularly tripod orientation calibration, masking by vessel noise, and the limited area monitored by two arrays. Future directions include deploying additional arrays, incorporating machine learning denoising (Miron et al. 2025), and integrating the system into broader conservation initiatives.

Conclusion

In this chapter authors provide an overview of three modules of a developing automated system that connects AI with PAM data for detecting, localizing and extracting real-time ecologically relevant information on SLEB.

First, the chapter introduced an automated pipeline capable of rapidly and accurately processing continuous PAM data to deliver reliable estimations on beluga acoustic presence and call-type counts, all while maintaining a minimal computational footprint suitable for embedded deployment in a future operational dynamic monitoring system.

Subsequently, the analysis of a dataset combining simultaneous visual counts and acoustic recordings was presented. It enabled testing the feasibility of automatically estimating local beluga abundance from acoustic activity. The results, however, show that such an approach still has important limitations. It remains very difficult to obtain precise beluga counts that fully cover the hydrophone's detection range. These counts can also contain errors, introducing label noise. Furthermore, the dataset size is relatively small, which limits the use of more advanced machine learning techniques that could capture temporal patterns and possibly integrate behavioral aspects. These issues will likely be even more pronounced when generalizing to new sites. Indeed, beluga behavior will differ across locations, affecting the relationship between vocal activity and abundance, and collecting accurate ground-truth data will be even harder in open or less constrained areas than BSM. To make progress with these objectives, future work should focus on methods to effectively count belugas in larger areas and obtain precise measurements of hydrophone detection ranges. Another promising venue to automatically estimate local beluga abundance could be to use precise localization methods to determine where belugas are vocalizing. By analyzing the geometry of detections and verifying whether multiple calls could physically come from the same individual, such approaches could provide more reliable minimum counts for local clusters of animals.

Finally, a TDOA-informed localization module was described. The algorithm shows promising performance in simulations and initial field tests. The agreement between theoretical and empirical error distributions provides encouraging evidence that the model captures the main sources of uncertainty affecting TDOA-based localization. Nevertheless, the absence of a true ground-truth reference for the in-situ detections prevents any definitive assessment of its validity at this stage. The upcoming controlled playback experiment, using an underwater loudspeaker emitting known signals at known positions and source levels, will constitute the first rigorous opportunity to test the system end-to-end—from signal detection and TDOA estimation to final 3D localization and uncertainty quantification. This experiment will determine whether the algorithm performs within its predicted error bounds and whether it can reliably reproduce known source coordinates under realistic noise conditions. Once such validation is completed, the proof of concept presented here will be ready for a real deployment with the goal of operationalizing a system dedicated to beluga conservation.

Acknowledgments The authors acknowledge the Réseau Québec Maritime for funding this project through the research program Programme de recherche sur l'Atténuation des impacts du transport maritime commercial et des activités portuaires sur les écosystèmes (PLAINE). They also thank Cristiane C.A. Martins (Parks Canada) and Valeria Vergara and her team (Raincoast Conservation Foundation) for providing data and research guidance that supported various components of the modules presented in this chapter.

Competing Interest Declaration The author(s) has no competing interests to declare that are relevant to the content of this manuscript.

References

Bell KL, Ephraim Y, Van Trees HL (1994) Explicit Ziv–Zakai lower bound for time delay estimation. IEEE Trans on Signal Process 42:3226–3230. https://doi.org/10.1109/78.330383

Cario G, Dabove P, Di Pietra V, Piras M, Lingua AM (2021) Accurate localization in acoustic underwater environments: a review. Sensors 21:762. https://doi.org/10.3390/s21030762

Chan YT, Ho KC (1994) A simple and efficient estimator for hyperbolic location. IEEE Trans on signal processing 42:1905–1915. https://doi.org/10.1109/78.301830

Chion C, Turgeon S, Cantin G, Michaud R, Ménard N, Lesage V, Parrott L, Beaufils P, Clermont Y, Gravel C (2018) A voluntary conservation agreement reduces the risks of lethal collisions between ships and whales in the St. Lawrence Estuary (Québec, Canada): from co-construction to monitoring compliance and assessing effectiveness. 13:e0202560. https://doi.org/10.1371/journal.pone.0202560

COSEWIC (2014) COSEWIC assessment and status report on the Beluga Whale Delphinapterus leucas, St. Lawrence Estuary population, in Canada. Committee on the Status of Endangered Wildlife in Canada. Ottawa. xii + 64 pp

Cotillard T, Sécheresse X, Aubin J, Mikus MA, Vergara V, Gambs S, Michaud R, Martins CC, Turgeon S, Chion C, Roca I (2024) Automatic detection and classification of beluga whale calls in the St. Lawrence estuary. The Journal of the Acoustical Society of America 156:3723–3740. https://doi.org/10.1121/10.0030472

DFO (2020) Action plan to reduce the impact of noise on the beluga whale and other marine mammals at risk in the St. Lawrence Estuary. Species at Risk act action plan series

DFO (2023) Abundance and population trajectory of St. Lawrence beluga

Fernandez E, Roca IT, Frizzle E, Michaud R, Mikus MA, Vergara V, Chion C (2025) Real-time acoustic monitoring: lightweight beluga call classification across habitats. Available at SSRN: https://ssrn.com/abstract=5451095

Gervaise C, Simard Y, Roy N, Kinda B, Menard N (2012) Shipping noise in whale habitat: Characteristics, sources, budget, and impact on belugas in Saguenay–St. Lawrence Marine Park hub. 132:76–89. https://doi.org/10.1121/1.4728190

Gillespie D, Mellinger DK, Gordon JO, Mclaren D, Redmond PA, McHugh R, Trinder PW, Deng XY, Thode A (2008) PAMGUARD: semiautomated, open-source software for real-time acoustic detection and localisation of cetaceans. Proceedings of the Institute of Acoustics 30:54–62. http://toc.proceedings.com/04143webtoc.pdf

Gillespie D, Palmer L, Macaulay J, Sparling C, Hastie G (2020) Passive acoustic methods for tracking the 3D movements of small cetaceans around marine structures. PLOS One 15: e0229058. https://doi.org/10.1371/journal.pone.0229058

Glotin H, Poupard M, Best P, Ferrari M (2021) Observations Stéréophoniques sur 4 ans de la Bouée BOMBYX au Sud du Parc National de Port-Cros: mégafaune et pollution anthropophonique. LOT 1: Le cas du cachalot. 24. ⟨hal-05010240⟩

Hyer MD, Anderson AT, Mann DA, Mooney TA, Aoki N, Jensen FH (2025) Robust real-time detection of right whale upcalls using neural networks on the edge. Ecological Informatics 103130. https://doi.org/10.1016/j.ecoinf.2025.103130

Lesage V (2021) The challenges of a small population exposed to multiple anthropogenic stressors and a changing climate: the St. Lawrence Estuary beluga. Polar Research 40. https://doi.org/10.33265/polar.v40.5523

Martin MJ, Halliday WD, Storrie L, Citta JJ, Dawson J, Hussey NE, Juanes F, Loseto LL, MacPhee SA, Moore L, Nicoll A (2023) Exposure and behavioral responses of tagged beluga whales (*Delphinapterus leucas*) to ships in the Pacific Arctic. Marine Mammal Science 39:387–421. https://doi.org/10.1139/as-2022-0052

Miron M, Keen S, Liu JY, Hoffman B, Hagiwara M, Pietquin O, Effenberger F, Cusimano M (2025) Biodenoising: animal vocalization denoising without access to clean data. arXiv preprint arXiv:2410.03427. https://doi.org/10.1109/ICASSP49660.2025.10889313

Roca IT, Kaleschke L, Van Opzeeland I (2023) Sea-ice anomalies affect the acoustic presence of Antarctic pinnipeds in breeding areas. Frontiers in Ecology and the Environment 21:227–233. https://doi.org/10.1002/fee.2622

Sjare BL, Smith T (1986) The vocal repertoire of white whales, *delphinapterus leucas*, summering in Cunningham inlet, northwest territories. Canadian Journal of Zoology 64:407–415

Small RJ, Brost B, Hooten M, Castellote M, Mondragon J (2017) Potential for spatial displacement of Cook Inlet beluga whales by anthropogenic noise in critical habitat. Endangered Species Research 32:43–57. https://doi.org/10.3354/esr00786

Spiesberger JL, Berchok C, Iyer P, Schoeny A, Sivakumar K, Woodrich D, Yang E, Zhu S (2021) Bounding the number of calling animals with passive acoustics and reliable locations. J Acoust Soc Am 150:1496–1504. https://doi.org/10.1121/10.0004994

Vergara V, Mikus MA (2019) Contact call diversity in natural beluga entrapments in an Arctic estuary: preliminary evidence of vocal signatures in wild belugas. Mar Mamm Sci 35:434–465. https://doi.org/10.1111/mms.12538

Vergara, V., Wood, J., Lesage, V., Ames, A., Mikus, M. A., & Michaud, R. (2021). Can you hear me? Impacts of underwater noise on communication space of adult, sub-adult and calf contact calls of endangered St. Lawrence belugas (Delphinapterus leucas). Polar Research, 40. https://doi.org/10.33265/polar.v40.5521

Vergara, V., Mikus, M. A., Chion, C., Lagrois, D., Marcoux, M., & Michaud, R. (2025). Effects of vessel noise on beluga (Delphinapterus leucas) call type use: ultrasonic communication as an adaptation to noisy environments?. Biology Open, 14(3):bio061783.